JOURNAL OF MULTIVARIATE ANALYSIS 7, 89–106 (1977)

Efficiency and Cramér–Rao Type Inequalities for Convex Loss Functions

ANDRZEJ KOZEK

Institute of Mathematics, Polish Academy of Sciences, 51-617 Wrocław, Kopernika 18, Poland

Communicated by K. Urbanik

A general method for obtaining inequalities of Cramér–Rao type for convex loss functions is presented. It is shown under rather weak assumptions that there are at least as many such inequalities as best unbiased estimators. More precisely, it is shown that an estimator is efficient with respect to an inequality of Cramér–Rao type if and only if it is the best in the class of unbiased estimators. Moreover, theorems of Blyth and Roberts ("Proceedings Sixth Berkeley Symposium on Math. Statist. Prob.,") and of Blyth (*Ann. Statist.* **2**, 464–473) are extended. We make an use of methods of convex analysis and properties of convex integral functionals on Orlicz spaces.

1. INTRODUCTION

In this paper we try to give a general and unified approach for obtaining inequalities of Cramér–Rao type for convex loss functions. At present there exists a large number of extensions of classical results of Rao (1945) and Cramér (1946), and we refer the reader to [8, Sect. 6] and [20, Chap. 4] for the literature (see also [5, 6, 13, 19]). In order to unify the different approaches Blyth and Roberts [3] introduced a general form of inequalities of Cramér–Rao type in the case of quadratic loss functions. According to their definition a lower bound for the risk functions is of Cramér–Rao type provided it is of a specified form and depends on estimators only through their expectations. On the other hand we have used in [8, Sect. 6] a Young–Fenchel inequality to obtain lower bounds for convex risk functions. Combining the above mentioned approaches we obtain in this paper a large class of Cramér–Rao type inequalities for convex loss functions. Moreover, we find some relations between Cramér–Rao type

Received September 1976.

AMS 1970 Subject Classification: Primary 62F99, 62F10; secondary 46E30, 46E40.

Key words: Conjugate convex function, convex loss function, Cramér–Rao type inequality, efficiency, lower bound for risk, Orlicz spaces, subdifferential, unbiased estimates.

ISSN 0047-259X

inequalities, an extension of the Lehmann–Scheffé lemma, and efficiencies of estimators with respect to Cramér–Rao type inequalities. Finally, we obtain that an estimator is the best unbiased one if and only if it is efficient with respect to an inequality of Cramér–Rao type.

The present paper gives, essentially, the proofs of the theorems announced in [10]. Moreover, there are results for Banach spaces with separable duals, whereas in [10] only a finite-dimensional case was considered.

To make the paper self-contained we collect in Section 2 the notions and theorems of convex analysis and Orlicz spaces used in Sections 3 and 4. In Section 3 we give a definition of Cramér–Rao type inequalities and, moreover, we prove several propositions. Propositions 3.4–3.6 extend theorems of Blyth and Roberts [3] and of Blyth [2]. In Theorems 1 and 2 in Section 4 we give the main results of the paper. These theorems establish an equality between the set of the best unbiased estimators and the set of estimators efficient with respect to Cramér–Rao type inequalities.

2. Preliminaries

2.1. Let $(T, \mathscr{A})$ be a measurable space and $\mathscr{P}$ a nonempty set of probability measures on $\mathscr{A}$. Let X be the set of decisions. We assume that X and its dual space Y are separable Banach spaces. Throughout the paper $\mathscr{B}_X$ and $\mathscr{B}_Y$ will stand for the σ-algebras of Borel subsets of X and Y, respectively.

2.2. A function f from $X \times T$ into $(-\infty, +\infty]$ is called a normal convex integrand if f is $\mathscr{B}_X \times \mathscr{A}$-measurable on $X \times T$ and for each $t \in T$ the function $f(\cdot, t)$ is convex, lower semicontinuous on X, and not identically $+\infty$.

The integrand g conjugate to f is a function from $Y \times T$ into $(-\infty, +\infty]$ given by

$$g(y, t) = \sup\{\langle x, y\rangle - f(x, t); x \in X\}, \tag{2.1}$$

where $\langle x, y\rangle$ stands for the value of $y \in Y$ at the point x.

The function g is a normal convex integrand on $Y \times T$ and f is in turn the integrand conjugate to g (see [12, 17]).

2.3. A normal convex integrand Φ defined on $X \times T$ is called an N-function if for every $t \in T$ the function $\Phi(\cdot, t)$ is continuous at the origin of X and satisfies the conditions

1. $\Phi(x, t) = \Phi(-x, t)$, $\Phi(0, t) \equiv 0$;

2. there exists $\alpha(t) \in (0, +\infty)$ and $\beta(t) \in (0, +\infty)$ such that $\Phi(x, t) \geqslant \alpha(t)$ provided $\| x \| \geqslant \beta(t)$.

The integrand Ψ defined on $Y \times T$ and conjugate to the N function Φ is an N-function, too. The symbols Φ and Ψ, with indices or not, will be used in the case of N-functions, only.

If $P \in \mathscr{P}$ is fixed, then a convex modular $I_{\Phi,P}$ given by

$$I_{\Phi,P}(x(\cdot)) = \int_T \Phi(x(t), t)\, P(dt) \tag{2.2}$$

is well defined and takes values in $[0, +\infty]$ when $x(\cdot)$ is an $\mathscr{A}$-measurable function. Since X is separable, both strong and weak measurabilities of $x(\cdot)$ are equivalent to $x^{-1}(B) \in \mathscr{A}$ for $B \in \mathscr{B}_X$ provided the σ-algebra $\mathscr{A}$ is P-complete.

Denote by $\mathscr{A}_P$ the completion of $\mathscr{A}$ with respect to P, the set of all $\mathscr{A}_P$-measurable functions from T into X by $\mathscr{L}_0(T, \mathscr{A}_P; X)$ and

$$\operatorname{dom} I_{\Phi,P} = \{x(\cdot) \in \mathscr{L}_0(T, \mathscr{A}_P; X): I_{\Phi,P}(x(\cdot)) < \infty\}.$$

The set $\operatorname{dom} I_{\Phi,P}$ is convex and it is called an Orlicz class with respect to the modular $I_{\Phi,P}$. Let $\mathscr{L}_{\Phi,P}$ be the linear hull spanned on $\operatorname{dom} I_{\Phi,P}$. Space $\mathscr{L}_{\Phi,P}$ is called an Orlicz space.

An Orlicz seminorm $N^1_{\Phi,P}$ is given on $\mathscr{L}_{\Phi,P}$ by

$$N^1_{\Phi,P}(x(\cdot)) = \inf\{(1/\xi)(1 + I_{\Phi,P}(\xi x(\cdot)));\ \xi > 0\},$$

whereas a Luxemburg seminorm $N^2_{\Phi,P}$ is defined by

$$N^2_{\Phi,P}(x(\cdot)) = \inf\{(1/\xi);\ I_{\Phi,P}(\xi x(\cdot)) \leqslant 1,\ \xi > 0\}, \qquad x(\cdot) \in \mathscr{L}_{\Phi,P}.$$

Both $N^1_{\Phi,P}$ and $N^2_{\Phi,P}$ are equivalent seminorms on $\mathscr{L}_{\Phi,P}$ because

$$N^2_{\Phi,P}(x(\cdot)) \leqslant N^1_{\Phi,P}(x(\cdot)) \leqslant 2N^2_{\Phi,P}(x(\cdot))$$

holds. $\mathscr{L}_{\Phi,P}$ endowed with the seminorm topology is a complete vector space. Moreover, if functions from $\mathscr{L}_{\Phi,P}$ which are equal P-a.e. are identified, then such a quotient space is denoted by $L_{\Phi,P}$, and $N^1_{\Phi,P}$ and $N^2_{\Phi,P}$ become norms on $L_{\Phi,P}$. Thus, $L_{\Phi,P}$ is a Banach space.

If Φ_1 and Φ_2 are N functions, then the following implications hold:

$$\begin{aligned} &\Phi_2(x, t) \leqslant \Phi_1(Kx, t) + h(t) \\ &\qquad \Rightarrow N^i_{\Phi_2,P}(x(\cdot)) \leqslant \text{const}\, N^i_{\Phi_1,P}(x(\cdot)) \qquad \text{for every } x(\cdot) \in \mathscr{L}_0(T, \mathscr{A}_P, X) \\ &\qquad \Rightarrow L_{\Phi_1,P} \subset L_{\Phi_2,P}. \end{aligned}$$

Here K is a constant and $h(\cdot)$ is a nonnegative P-summable function. When X is separable and P is purely nonatomic all these conditions are equivalent.

Let Ψ be the conjugate of Φ. Since Ψ is an N-function the Orlicz spaces $\mathscr{L}_{\Psi,P}$, $L_{\Psi,P}$ and norms $N^i_{\Psi,P}$, $i = 1, 2$, are defined as above but with the use of Ψ and Y instead of Φ and X. Clearly $L_{\Psi,P}$ is a Banach space.

Let $\{T_i\}_{i=1}^{\infty}$ be an increasing sequence of sets such that $T_i \in \mathscr{A}$ and $P(T \setminus \bigcup_{i=1}^{\infty} T_i) = 0$. The function Φ satisfies so called Condition B with respect to $\{T_i\}$ if functions $f_n(\cdot)$ given by

$$f_n(t) = \sup\{\Phi(x, t); \| x \| \leqslant n\}, \qquad n \in \mathbf{N} \tag{2.3}$$

are P-summable on each of the sets T_i, $i \in \mathbf{N}$.

Since X is separable $f_n(\cdot)$ is $\mathscr{A}_P$-measurable for each $n \in \mathbf{N}$. Thus, in Condition B only the summability of f_n on T_i is assumed. In the sequel we will assume that Φ satisfies Condition B. This implies that $E_{\Phi,P}$, the largest closed subspace of $\operatorname{dom} I_{\Phi,P}$, is a nontrivial subspace and contains all bounded $\mathscr{A}_P$-measurable functions with supports in T_i, $i \in \mathbf{N}$. Conversely, if $E_{\Phi,P}$ contains all bounded $\mathscr{A}_P$-measurable functions with supports in T_i, $i \in \mathbf{N}$, then Condition B is fulfilled provided X is separable and P has no atoms.

Let $d(x(\cdot), E_{\Phi,P})$ denote the distance in the $N^1_{\Phi,P}$ norm of $x(\cdot) \in L_{\Phi,P}$ from $E_{\Phi,P}$ and let $\Pi(E_{\Phi,P}, 1)$ be the set of all $x(\cdot) \in L_{\Phi,P}$ for which $d(x(\cdot), E_{\Phi,P}) < 1$. Then $\operatorname{intdom} I_{\Phi,P} = \Pi(E_{\Phi,P}, 1)$.

The sets T_i in Condition B have an auxiliary character and neither $L_{\Phi,P}$ nor $E_{\Phi,P}$ depend on the choice of $\{T_i\}$. None the less, the functions f_n given by (2.3) need not satisfy Condition B for every sequence $\{T_i\}$ such that $T_{i+1} \supset T_i$ and $P(T \setminus \bigcup T_i) = 0$.

If $E_{\Phi,P}$ is endowed with the $N^i_{\Phi,P}$-norm topology, then the dual space of $E_{\Phi,P}$ is isometrically isomorphic with the Orlicz space $L_{\Psi,P}$ endowed with the $N^j_{\Psi,P}$-norm topology, $i \neq j$, $i, j = 1, 2$. Moreover, the formula

$$\varphi(x(\cdot)) = \langle x(\cdot), y(\cdot) \rangle_P \tag{2.4}$$

yields a univocal representation of linear continuous functionals $\varphi \in E'_{\Phi,P}$, where

$$\langle x(\cdot), y(\cdot) \rangle_P = \int_T \langle x(t), y(t) \rangle \, P(dt).$$

In the sequel it will be convenient to identify the elements of $L_{\Psi,P}$ with the corresponding elements of $E'_{\Phi,P}$ (see [7, 9]). The correspondence is given by formula (2.4).

2.4. The N-function Φ satisfies so called Condition Δ_2 if

$$\Phi(2x, t) \leqslant K\Phi(x, t) + h(t)$$

holds for some positive constant K and some P-summable function h. In this case $E_{\Phi,P} = L_{\Phi,P}$ and $\operatorname{dom} I_{\Phi,P} = L_{\Phi,P}$.

Conversely, if X is separable and P is purely nonatomic, then $E_{\Phi,P} = L_{\Phi,P}$ implies that Φ satisfies Condition Δ_2 (see [9]).

EXAMPLE. Let $\Phi(x, t) = (1/p)\| x \|^p$, $p \in [1, \infty)$. In case $p > 1$ we have $\Psi(y, t) = \frac{1}{q}\| y \|^q$, where $(1/p) + (1/q) = 1$ and in case $p = 1$

$$\Psi(y, t) = \begin{cases} 0, & \text{if } \| y \| \leqslant 1, \\ +\infty, & \text{if } \| y \| > 1. \end{cases}$$

The Orlicz space $L_{\Phi,P}$ is here the usual space $L^X_{p,P}$, $E_{\Phi,P} = L_{\Phi,P}$ because Φ satisfies Condition Δ_2 and the norms $N^i_{\Phi,P}$, $i = 1, 2$, differ from the usual $L^X_{p,P}$-norms with constant multipliers only (the multipliers are even equal to one when $p = 1$). Moreover, in this case $L_{\Psi,P} = L^Y_{q,P}$ and hence $L^Y_{q,P}$ may be considered as the dual of $L^X_{p,P}$ (see [8] and a paper of L.Schwartz in "1974–1975 Seminaire Maurey–Schwartz").

2.5. $L_{\Psi,P}$ is a closed subspace of $L'_{\Phi,P}$. The dual space of $L_{\Phi,P}$ is a sum $L_{\Psi,P} \oplus \Lambda$, where Λ is a closed subspace of $L'_{\Phi,P}$ consisting of all functionals $\varphi \in L'_{\Phi,P}$ which vanish on $E_{\Phi,P}$. The elements of Λ are called singular functionals. If Φ satisfies Condition Δ_2, then $\Lambda = 0$.

Let f be a normal convex integrand, and let I_f, given by

$$I_{f,P}(x(\cdot)) = \int_T f(x(t), t)\, P(dt), \tag{2.5}$$

be well defined on $L_{\Phi,P}$ and take values in $(-\infty, +\infty]$.

The subdifferential of $I_{f,P}$ at a point $x_0(\cdot) \in L_{\Phi,P}$ such that $x_0(\cdot) \in \operatorname{dom} I_{f,P}$ is denoted $\partial I_{f,P}(x_0(\cdot))$ and consists of all functionals $\varphi \in L'_{\Phi,P}$ satisfying for all $x(\cdot) \in L_{\Phi,P}$ the inequality

$$I_{f,P}(x(\cdot)) \geqslant I_{f,P}(x(\cdot)) + \langle x(\cdot) - x_0(\cdot), \varphi\rangle. \tag{2.6}$$

The symbol $\langle x(\cdot), \varphi\rangle$ denotes the value of φ at $x_0(\cdot)$. The elements of $\partial I_{f,P}(x_0(\cdot))$ are called subgradients of $I_{f,P}$ at $x_0(\cdot)$.

The subdifferential $\partial f(x, t)$ is defined for $x_0 \in \operatorname{dom} f(\cdot, t)$ and consists of all functionals $y \in Y$ such that

$$f(x, t) \geqslant f(x_0, t) + \langle x - x_0, y\rangle$$

holds for all $x \in X$. Let $D_{f,P}(x_0(\cdot))$ stand for the set of all fundtions $y(\cdot) \in L_{\Psi,P}$ such that $y(t) \in \partial f(x_0(t), t)$ P-a.e. Let us denote by $K_{f,P}(x_0(\cdot))$ the set of all functionals $\varphi \in L'_{\Phi,P}$ such that

$$\langle x(\cdot) - x_0(\cdot), \varphi\rangle \leqslant 0$$

for every $x(\cdot) \in \operatorname{dom} I_{f,P}$ holds.

The subdifferential $\partial I_{f,P}(x_0(\cdot))$ admits a representation

$$\partial I_{f,P}(x_0(\cdot)) = D_{f,P}(x_0(\cdot)) + K_{f,P}(x_0(\cdot)) \tag{2.7}$$

and every element of $K_{f,P}(x_0(\cdot))$ is a singular functional from Λ. If $x_0(\cdot) \in \operatorname{int} \operatorname{dom} I_{f,P}(x_0(\cdot))$, then $K_{f,P}(x_0(\cdot)) = 0$ (see [9, 12]).

2.6. If $I_{f,P}$ is finite and continuous at $x_0(\cdot)$, then it is subdifferentiable, i.e., $\partial I_{f,P}(x_0(\cdot)) \neq \varnothing$. The continuity of $I_{f,P}$ at $x_0(\cdot)$ is equivalent to the boundedness from above of $I_{f,P}$ on a neighborhood of $x_0(\cdot)$. Moreover, $I_{f,P}$ is continuous on int dom $I_{f,P}$. If $I_{f,p}$ is finite at every point of $L_{\Phi,P}$, then it is continuous at every point of $L_{\Phi,P}$.

2.7. Let U and W be vector spaces paired with respect to a bilinear form $\langle \cdot, \cdot \rangle$ on $U \times W$. If $r(\cdot)$ is a convex function from U into $(-\infty, +\infty]$, then the conjugate of $r(\cdot)$ given by

$$m(w) = \sup\{\langle u, w \rangle - r(u);\ u \in U\}$$

is a $\sigma(W, U)$-lower semicontinuous convex function from W into $(-\infty, +\infty]$. Function $r(\cdot)$ is the conjugate of $m(\cdot)$ if and only if it is $\sigma(U, W)$-lower semicontinuous. A subgradient of $r(\cdot)$ at $u_0 \in \operatorname{dom} r(\cdot)$ is a vector $w_0 \in W$ such that

$$r(u) \geqslant r(u_0) + \langle u - u_0, w_0 \rangle$$

holds for every $u \in U$. The set $\partial r(u_0)$ of all subgradients of $r(\cdot)$ at u_0 is called a subdifferential of r at u_0.

Moreover, if $r(\cdot)$ and $m(\cdot)$ are conjugate to each other, then

$$w \in \partial r(u) \Leftrightarrow u \in \partial m(w) \Leftrightarrow r(u) = \langle u, w \rangle - m(w) \tag{2.8}$$

(see [4, 13]).

If N is a linear manifold in U, r and m are conjugate to each other, r is finite and continuous at $u_0 \in N$, then r attains at u_0 its infimum over N if and only if there exists an element $w_0 \in \partial r(u_0)$ such that for every $u \in N$

$$\langle u - u_0, w_0 \rangle = 0$$

holds (see [8]).

3. A General Form of Inequalities of Cramér–Rao Type

Let T, $\mathscr{A}$, $\mathscr{P}$, X, and Y be as in Section 2.1. We assume throughout the paper that the loss function $L(x, t, P)$ is for every $P \in \mathscr{P}$ a nonnegative normal convex integrand on $X \times T$ such that for every $P \in \mathscr{P}$ the following conditions are fulfilled:

1. $L(0, t, P)$ is a summable function;
2. functions $\bar{L}(\alpha, t, P)$ given by

$$\bar{L}(\alpha, t, P) = \sup\{L(x, t, P); \| x \| \leqslant \alpha\}$$

are finite P-a.e. for every real number α;

3. there exists a P-summable function $h_P(\cdot)$ such that

$$\| x \| \leqslant K_P \max\{L(x, t, P), L(-x, t, P)\} + h_P(t)$$

holds, where K_P is a constant.

PROPOSITION 3.1. *Let L be a loss function satisfying conditions given above. Let $\Phi_P(x, t)$ be given by*

$$\Phi_P(x, t) = \max\{L(x, t, P), L(-x, t, P)\} - L(0, t, P). \tag{3.1}$$

Then Φ_P is an N-function, there exists sets $T_{P,n}$ such that Φ_P satisfies Condition B *with respect to $\{T_{P,n}\}$, $L_{\Phi,P} \subset L^X_{1,P}$ and the imbedding of $L_{\Phi,P}$ into $L^X_{1,P}$ is continuous.*

Proof. Clearly Φ_P given by (3.1) is a nonnegative normal convex integrand such that $\Phi_P(x, t) = \Phi_P(-x, t)$ and $\Phi_P(0, t) \equiv 0$. Condition 2 implies that $\Phi_P(\cdot, t)$ is finite on X and hence continuous for P-a.e. $t \in T$. Condition 3 yields the inequality

$$\| x \| \leqslant \Phi_P(K_P x, t) + h_P'(t), \tag{3.2}$$

where $h_P'(\cdot)$ is a P-summable function. Now, by the definition of conjugate function we obtain that Ψ_P, the conjugate of Φ_P, satisfies the inequality

$$\Psi_P(y, t) \leqslant \delta_{K(0,1)}(K_p y) + h_p'(t),$$

where

$$\delta_{K(0,1)}(y) = \begin{cases} 0, & \text{if } \| y \| \leqslant 1, \\ +\infty & \text{elsewhere.} \end{cases}$$

Thus, Ψ_P is continuous at zero and this implies the existence of $\alpha(t) \in (0, \infty)$ and $\beta(t) \in (0, \infty)$ such that

$$\Phi_P(x, t) \geqslant \alpha(t) \qquad \text{whenever} \quad \| x \| \geqslant \beta(t)$$

(see the proof of [7, Proposition 4.6] for the last implication). Thus Φ_P is an N-function. The inclusion $L_{\Phi,P} \subset L^X_{1,P}$ and the continuity of the imbedding of $L_{\Phi,P}$ into $L^X_{1,P}$ follow from implications given in Section 2.3 (cf. [9, Theorem 1.8].)

Finally we prove that Φ_P satisfies Condition B with respect to sets $\{T_{P,n}\}$ given by

$$T_{P,n} = T \setminus \bigcup_{i=n}^{\infty} A_{P,i},$$

where

$$A_{P,i} = \{t \in T: \bar{L}(i, t, P) > m(i)\}$$

and where the number $m(i)$ satisfies the condition $P(A_{P,i}) < 2^{-i}$, $i \in \mathbf{N}$. If $j < i$, then $t \in T_{P,i}$ implies $t \notin A_{P,i}$, $\bar{L}(i, t, P) \leqslant m(i)$ and hence $\bar{L}(j, t, P) \leqslant m(i)$. If $j \geqslant i$ and $t \in T_{P,i}$, then $t \notin A_{P,j}$ and $\bar{L}(j, t, P) \leqslant m(j)$. Thus for every j and i $\bar{L}(j, \cdot, P)$ is bounded on T_i, proving the assertion.

Remark 1. It is easy to see that if X is a finite-dimensional space and $L(x, t, P)$ is finite and convex for every t and P, then Condition 2 on the loss function is fulfilled.

Remark 2. Let us point out that Condition 2 on the loss function L yields that Φ_P given by (3.1) fulfills Condition B. Condition 3 implies the existence of functions $\alpha(t)$ and $\beta(t)$ from the definition of N-function and also the inclusion $L_{\Phi,P} \subset L^X_{1,P}$. Hence a Bochner integral $E_P x(\cdot)$ is well defined for $x(\cdot) \in \bigcap_{P \in \mathscr{P}} \mathscr{L}_{\Phi,P}$.

One can give a weaker condition than Condition 3. However, in the case of P purely nonatomic both these conditions 3 and 3′ given below are equivalent.

CONDITION 3′. There exists a positive P-summable function $\alpha_P(\cdot)$ and a constant C_P such that if Φ_P is given by (3.1), then

$$\Phi_P(C_P \alpha_P(t) x / \| x \|, t) \geqslant \alpha_P(t).$$

It is easy to see that if we put $\beta_P(t) = C_P \alpha_P(t)$, then Condition 3′ yields

$$\| x \| \geqslant \beta_P(t) \Rightarrow \Phi_P(x, t) \geqslant \alpha_P(t).$$

From the proof of [7, Proposition 4.6] it follows that if $\| y \| \leqslant C_P$ and Ψ_P is the conjugate of Φ_P, then $\Psi_P(y, t) \leqslant \frac{1}{2}\alpha_P(t)$. Hence $L_{\Psi,P} \supset L^X_{\infty,P}$. Since

$$N^1_{\Phi,P}(x(\cdot)) = \sup\{\langle x(\cdot), y(\cdot)\rangle_P ; I_{\Psi,P}(y(\cdot)) \leqslant 1\}$$

and since $\alpha_P(\cdot)$ is P-summable we obtain

$$\begin{aligned} N^1_{\Phi,P}(x(\cdot)) &\geqslant \text{const} \sup\{\langle x(\cdot), y(\cdot)\rangle_P ; \sup \operatorname{ess}_P \| y(\cdot) \| \leqslant 1\} \\ &= \text{const} \| x(\cdot) \|_{L^X_{1,P}}. \end{aligned}$$

Let $\mathscr{E}$ be the set of estimators under consideration. Throughout the paper we assume that

$$\mathscr{E} = \bigcap_{P \in \mathscr{P}} \mathscr{L}_{\Phi,P} \cap \mathscr{L}_0(T, \mathscr{A}_{\mathscr{P}}\, ; X),$$

where $\mathscr{A}_{\mathscr{P}}$ is the smallest σ-algebra containing $\mathscr{A}$ and all subsets of those sets from $\mathscr{A}$ which have P-measure zero for every $P \in \mathscr{P}$. $\mathscr{L}_0(T, \mathscr{A}_{\mathscr{P}}; X)$ stands here for the set of all $\mathscr{A}_{\mathscr{P}}$-measurable functions.

If $x(\cdot)$ is an estimator and if $P \in \mathscr{P}$, then the risk function is given by

$$R(x(\cdot), P) = \int_T L(x(t), t, P)\, P(dt), \qquad x(\cdot) \in \mathscr{E}. \tag{3.2'}$$

As is easy to see, $\mathscr{E}$ is the set of all $\mathscr{A}_{\mathscr{P}}$-measurable functions $x(\cdot)$ for which there exists a positive number $\alpha = \alpha_P$ such that both $R(\alpha x(\cdot), P)$ and $R(-\alpha x(\cdot), P)$ are finite for every $P \in \mathscr{P}$.

We have defined the risk function as a function over $\mathscr{E} \times \mathscr{P}$. However, if $P \in \mathscr{P}$ is fixed, then we may and in fact we will consider $R(\cdot, P)$ given by (3.2′) as a convex functional on $L_{\Phi,P}$. Another symbol which has more than one meaning is $x(\cdot)$. If $x(\cdot) \in L_{\Phi,P}$, then it is a class of functions $\mathscr{A}_P$-measurable and equal P-a.e. to each other. If $x(\cdot) \in \mathscr{E}$, then it is an $\mathscr{A}_{\mathscr{P}}$-measurable function. Sylbol $x(\cdot)$ will also be used as a class of $\mathscr{A}_{\mathscr{P}}$-measurable functions which are equal P-a.e. to $x(\cdot)$ for every $P \in \mathscr{P}$. However, these ambiguities simplify notation and do not lead to misunderstandings.

Let $M(y, t, P)$ stand for the function conjugate to the loss $L(x, t, P)$, $P \in \mathscr{P}$. $M(y, t, P)$ is given by

$$M(y, t, P) = \sup\{\langle x, y\rangle - L(x, t, P);\, x \in X\}. \tag{3.3}$$

Let $I_M(y(\cdot), P)$ be a convex integral functional defined by

$$I_M(y(\cdot), P) = \int_T M(y(t), t, P)\, P(dt) \tag{3.4}$$

for $y(\cdot) \in \mathscr{L}_{\Psi,P}$.

In the sequel we will make use of the properties of convex functionals $R(\cdot, P)$ and $I_M(\cdot, P)$ which are given in the following two lemmas. We assume henceforth that Φ_P is related to the loss function by formula (3.1) and we will use the notations given in Section 2. For every $P \in \mathscr{P}$ a pairing between $L_{\Phi,P}$ and $L_{\Psi,P}$ is established by

$$\langle x(\cdot), y(\cdot)\rangle_P = \int_T \langle x(t), y(t)\rangle\, P(dt) \qquad (= E_P\langle x(\cdot), y(\cdot)\rangle). \tag{3.5}$$

LEMMA 3.2. *Let $P \in \mathscr{P}$. The convex integral functionals $R(\cdot, P)$ and $I_M(\cdot, P)$ defined on $L_{\Phi,P}$ and $L_{\Psi,P}$, respectively, are conjugate to each other; i.e.,*

$$R(x(\cdot), P) = \sup\{\langle x(\cdot), y(\cdot)\rangle_P - I_M(y(\cdot), P);\, y(\cdot) \in L_{\Phi,P}\} \tag{3.6}$$

holds for every $x(\cdot) \in L_{\Phi,P}$, and

$$I_M(y(\cdot), P) = \sup\{\langle x(\cdot), y(\cdot)\rangle_P - R(x(\cdot), P);\, x(\cdot) \in L_{\Phi,P}\} \tag{3.7}$$

holds for every $y(\cdot) \in L_{\Psi,P}$.

Proof. In view of Proposition 1.3 and [7, Theorem 1.2] formula (3.6) holds provided the following two conditions are fulfilled: (1) every $\mathscr{A}_P$-measurable function $y(\cdot)$ such that $I_M(y(\cdot), P)$ is finite is an element of $L_{\Psi,P}$ and (2) there exists $y(\cdot) \in \mathscr{L}_0(T, \mathscr{A}_P; Y)$ such that $I_M(y(\cdot), P)$ is finite.

Formula (3.1) yields

$$L(x, t, P) - L(0, t, P) \leqslant \Phi_P(x, t). \tag{3.8}$$

The definition of conjugate function implies that

$$\Psi_P(y, t) \leqslant M(y, t, P) + L(0, t, P).$$

Hence P-summability of $M(y(t), t, P)$ implies P summability of $\Psi_P(y(t), t)$. Then, by the definition of an Orlicz space (see Section 2.3) $y(\cdot) \in L_{\Psi,P}$. Moreover, (3.3) yields the inequality

$$-L(0, t, P) \leqslant M(0, t, P) \leqslant 0.$$

This means that if $y_0(t) \equiv 0$, then $I_M(y_0(\cdot), P)$ is finite. Thus (3.6) holds. The loss function $L(x, t, P)$ is finite P-a.e. and by (3.8) the convex functional $R(\cdot, P)$ is finite at the point $x_0(t) \equiv 0$. Moreover, functions of the form $x \cdot \chi_D(\cdot)$ are elements of $L_{\Phi,P}$ whenever D is a measurable subset of some $T_{P,n}$. Hence [7, Proposition 1.4 and Theorem 1.2] imply (3.7).

LEMMA 3.3. *If $P \in \mathscr{P}$ and a pairing $\langle \cdot, \cdot \rangle_P$ between $L_{\Phi,P}$ and $L_{\Psi,P}$ is given by* 3.5, *then $(L_{\Phi,P}, L_{\Psi,P})$ is a dual pair.*

Proof. $L_{\Phi,P}$ contains all functions of the form $x \cdot \chi_D(\cdot)$, where $D \in \mathscr{A}_P$, $D \subset T_{P,n}$. Therefore the equality

$$\langle x \cdot \chi_D(\cdot), y(\cdot)\rangle_P = \int_D \langle x, y(t)\rangle\, P(dt) = 0$$

for each $D \in \mathscr{A}_P$, $D \subset T_{P,n}$, $n \in \mathbf{N}$, implies $\langle x, y(t)\rangle = 0$ P-a.e. By the separability of X we obtain that $y(t) = 0$ P-a.e.

If we take $L_1(x, t, P) = \Phi_P(x, t)$, then the function $\Phi_P{}^1(x, t, P)$ given by (3.1) with $L(x, t, P)$ substituted by $L_1(x, t, P)$ is equal to $\Phi_P(x, t)$. Thus (3.6) yields

$$I_{\Phi,P}(x(\cdot)) = \sup\{\langle x(\cdot), y(\cdot)\rangle_P - I_{\Psi,P}(y(\cdot)); y(\cdot) \in L_{\Psi,P}\}, \tag{3.9}$$

where $I_{\Phi,P}$ is given by (2.2) and $I_{\Psi,P}$ is given by

$$I_{\Psi,P}(y(\cdot)) = \int_T \Psi_P(y(t), t)\, P(dt). \tag{3.10}$$

Suppose that $\langle x(\cdot), y(\cdot)\rangle_P = 0$ holds for every $y(\cdot) \in L_{\Psi,P}$. Since $\Psi_P(y, t)$ is an N-function we obtain from (3.9) that $I_{\Phi,P}(\xi x(\cdot)) = 0$ for every $\xi > 0$. Hence $N^1_{\Phi,P}(x(\cdot)) = 0$; i.e., $x(t) = 0$ P-a.e.

Lemmas 3.2 and 3.3 imply immediately the following corollary.

COROLLARY. *For every $P \in \mathscr{P}$ convex integral functionals $R(\cdot, P)$ and $I_M(\cdot, P)$ are lower semicontinuous in topologies $\sigma(L_{\Phi,P}, L_{\Psi,P})$ and $\sigma(L_{\Psi,P}, L_{\Phi,P})$, respectively.*

Since for every $P \in \mathscr{P}$ Orlicz spaces $\mathscr{L}_{\Phi,P}$ contain the set of estimators $\mathscr{E}$, formula (3.6) admits the following interpretation. Whatever $x(\cdot) \in \mathscr{E}$ and $y_P(\cdot) \in \mathscr{L}_{\Psi,P}$, the inequality

$$R(x(\cdot), P) \geqslant E_P\langle x(\cdot), y_P(\cdot)\rangle - I_M(y_P(\cdot)) \tag{3.11}$$

holds for all $P \in \mathscr{P}$. Moreover, if $x(\cdot) \in \mathscr{E}$ is fixed, then $R(x(\cdot), P)$ is the supremum of the right-hand side of (3.11) over $y_P(\cdot) \in \mathscr{L}_{\Psi,P}$. The right-hand side of (3.11) appears to give a lower bound for the risk of $x(\cdot)$. However, such a bound is useless because it depends on $x(\cdot)$ rather than (e.g.) on the expectation of $x(\cdot)$. Moreover, rather than compute $E_P\langle x(\cdot), y_P(\cdot)\rangle$ and $I_M(y_P(\cdot))$ it would be more convenient to compute the risk $R(x(\cdot), P)$ itself. In general, a lower bound for a risk function R defined on $\mathscr{E} \times \mathscr{P}$ is a function B from $\mathscr{E} \times \mathscr{P}$ into $[0, +\infty]$ such that the inequality

$$R(x(\cdot), P) \geqslant B(x(\cdot), P)$$

holds for every $(x(\cdot), P) \in \mathscr{E} \times \mathscr{P}$. The "useful" lower bounds should depend on $x(\cdot) \in \mathscr{E}$ only through expectation of $x(\cdot)$. This means that the funcion B should be of the form

$$B(x(\cdot), P) = K(E_P x(\cdot), P), \tag{3.12}$$

where K is a function from $X \times \mathscr{P}$ into $[0, +\infty]$ and $E_P x(\cdot)$ is a Bochner integral. The expectation $E_P x(\cdot)$ is well defined because, by Proposition 3.1, $L_{\Phi,P} \subset L^X_{1,P}$.

DEFINITION. If the right-hand side of inequality (3.11) depends on $x(\cdot)$ only through expectation $E_P x(\cdot)$, then it is called a lower bound of Cramér–Rao type for the risk R and (3.11) is called an inequality of Cramér–Rao type.

Clearly, (3.11) is an inequality of Cramér–Rao type if and only if $E_P\langle x(\cdot), y_P(\cdot)\rangle$ depends on $x(\cdot)$ only through expectation $E_P x(\cdot)$. The following lemma characterizes functions $y_P(\cdot) \in \mathscr{L}_{\Psi,P}$, $P \in \mathscr{P}$, which lead to Cramér–Rao type inequalities.

LEMMA 3.4. *Let $y_P(\cdot) \in \mathscr{L}_{\Psi,P}$ for every $P \in \mathscr{P}$. Then*

$$E_P x_1(\cdot) \overset{\mathscr{P}}{\equiv} E_P x_2(\cdot) \Rightarrow E_P\langle x_1(\cdot), y_P(\cdot)\rangle \overset{\mathscr{P}}{\equiv} E_P\langle x_2(\cdot), y_P(\cdot)\rangle \tag{3.13}$$

if and only if

$$E_P\langle x(\cdot), y_P(\cdot)\rangle \overset{\mathscr{P}}{\equiv} 0 \tag{3.14}$$

for each $x(\cdot) \in \mathscr{E}$ such that $E_P x(\cdot) \equiv^{\mathscr{P}} 0$.

Proof. The proof of this lemma is obvious.

It is remarkable that functions $y_P(\cdot)$, $P \in \mathscr{P}$, leading to Cramér–Rao type inequalities are rather insensitive to a change of loss function from L into, say L'. The only requirement is that

$$\mathscr{L}_{\phi,P} = \mathscr{L}_{\phi',P} \qquad \text{for every } P \in \mathscr{P}.$$

Note that if inequality (3.11) is of Cramér–Rao type, then it may be rewritten as

$$R(x(\cdot), P) \geqslant b(E_P x(\cdot)) + d(P), \tag{3.15}$$

where

$$b(E_P x(\cdot)) = E_P\langle x(\cdot), y_P(\cdot)\rangle \qquad \text{and} \qquad d(P) = -I_M(y_P(\cdot), P).$$

By Lemma 3.4 functions $y_P(\cdot)$, $P \in \mathscr{P}$, fulfil condition (3.14) if and only if functions $cy_P(\cdot)$, $P \in \mathscr{P}$, fulfil this condition. Thus, for every $c \in R$, we have

$$R(x(\cdot), P) \geqslant cE_P\langle x(\cdot), y_P(\cdot)\rangle - I_M(cy_P(\cdot), P). \tag{3.16}$$

The function $v_P{}^*$ defined on R by

$$v_P{}^*(c) = I_M(cy_P(\cdot), P) = \int M(cy_P(t), t, P)\, P(dt)$$

is lower semicontinuous and convex because $M(\cdot, t, P)$ is convex and lower semicontinuous. Taking supremum over $c \in R$ we obtain from (3.16)

$$R(x(\cdot), P) \geqslant v_P(E_P\langle x(\cdot), y_P(\cdot)\rangle) = v_P(b(E_P x(\cdot))), \tag{3.17}$$

where v_P is the conjugate of $v_P{}^*$. Inequality (3.17) is of the form given by (3.12) but it may happen that it is not of Cramér–Rao type. However, if for every $P \in \mathscr{P}$ the functions v_P are finite (thus continuous), then v_P are subdifferentiable and hence there exist c_P such that

$$v_P(b(E_P x(\cdot))) = c_P b(E_P x(\cdot)) - v_P{}^*(c_P).$$

This means that (3.17) is of Cramér–Rao type, provided v_P are finite. Note that (3.17) is useful for obtaining various nonstandard Cramér–Rao type inequalities (cf. [8, Sect. 6])

Propositions 3.5 and 3.6 given below extend Theorem 2 of Blyth and Roberts [3]. Let S stand for the intersection of all sufficient σ-algebras S_β containing all $\mathscr{P}$-null sets.

PROPOSITION 3.5. *Let $L(\cdot, \cdot, P)$ be $\mathscr{B}_X x S$-measurable for every $P \in \mathscr{P}$ and let $\bar{L}(\alpha, \cdot, P)$ be P-summable for some $\alpha_P > 0$. If condition* (3.14) *is fulfilled for every $x(\cdot) \in \mathscr{E}$ such that $E_P x(\cdot) \overset{\mathscr{P}}{=} 0$, then $y_P(\cdot)$, $P \in \mathscr{P}$, is S-measurable for every $P \in \mathscr{P}$.*

LEMMA 3.5.1. *Let $L(\cdot, \cdot, P)$ be $\mathscr{B}_X x S$-measurable for every $P \in \mathscr{P}$ and let $\bar{L}(\alpha, \cdot, P)$ be P-summable for some $\alpha_P > 0$. If $x(\cdot) \in L_{\Phi,P}$, $y_P(\cdot) \in L_{\Psi,P}$, and S is a sub-σ-algebra of $\mathscr{A}$, then the conditional expectations $E_P(x(\cdot) \mid S)$ and $E_P(y_P(\cdot) \mid S)$ are well defined, they are elements of $L_{\Phi,P}$ and $L_{\Psi,P}$, respectively, and the equalities*

$$\begin{aligned} E_P\langle E_P(x(\cdot) \mid S), y_P(\cdot)\rangle &= E_P\langle E_P(x(\cdot) \mid S), E_P(y_P(\cdot) \mid S)\rangle \\ &= E_P\langle x(\cdot), E_P(y_P(\cdot) \mid S)\rangle \end{aligned} \tag{3.18}$$

hold.

Proof. By the Radon–Nikodym theorem for Banach space valued measures [15] there exists an S-measurable function $E_P(x(\cdot) \mid S)$ from T into X such that

$$\int_A x(t)\, P(dt) = \int_A E_P(x(\cdot) \mid S)(t)\, P(dt)$$

holds for every $A \in S$. As in the finite-dimensional case, $E_P(x(\cdot) \mid S)$ is called the conditional expectation of $x(\cdot)$ with respect to S.

Similarly the conditional expectation $E_P(y_P(\cdot) \mid S)$ is well defined provided $E_P y_P(\cdot)$ exists. The existence of $E_P y_P(\cdot)$ is guaranteed by the P summability of $\bar{L}(\alpha_P, \cdot, P)$. Indeed, if $\text{supess}_P \| x(\cdot)\| \leqslant \alpha_P$, then $I_{\Phi,P}(x(\cdot)) \leqslant c_P$. By convexity of Φ_P this yields $I_{\Phi,P}(\min(1, 1/c_P)\, x(\cdot)) \leqslant 1$, i.e., $N^2_{\Phi,P}(x(\cdot)) \leqslant$ const. Since

$$N^1_{\Psi,P}(y(\cdot)) = \sup\{\langle x(\cdot), y(\cdot)\rangle_P \,;\, N^2_{\Phi,P}(x(\cdot)) \leqslant 1\}$$

(see [7, Sect. 1]) we obtain that $N^1_{\Psi,P}(y(\cdot)) \geqslant$ const $N^Y_{1,P}(y(\cdot))$, where $N^Y_{1,P}(y(\cdot))$ denotes the usual norm in $L^Y_{1,P}$.

Equalities (3.18) can be proved by standard methods. We note that

$$\langle E_P(x(\cdot) \mid S), E_P(y_P(\cdot) \mid S)\rangle = E_P(\langle E_P(x(\cdot) \mid S), y_P(\cdot)\rangle \mid S) \tag{3.19}$$

holds when $y_P(\cdot)$ is a step function. In a general case $y_P(\cdot)$ can be approximated by step functions in such a way that (3.19) follows from the dominated convergence theorem and from the definition of conditional expectation. This yields the first equality in (3.18). Similarly the second equality can be proved.

$E_P(x(\cdot) \mid S)$ (resp. $E_P(y_P(\cdot) \mid S)$) is an element of $L_{\Phi,P}$ (resp. $L_{\Psi,P}$) provided $x(\cdot) \in L_{\Phi,P}$ (resp. $y_P(\cdot) \in L_{\Psi,P}$). This follows from the inequalities

$$(\Phi_P(E_P(x(\cdot) \mid S), \cdot, P) \leqslant E_P(\Phi_P(x(\cdot), \cdot, P) \mid S)$$

(resp. $\Psi_P(E_P(y_P(\cdot) \mid S), \cdot, P) \leqslant E_P(\Psi_P(y_P(\cdot), \cdot, P) \mid S)$), which are part icular cases of an extended version of Jensen's inequality [18] (see also [1] in the case of separable, reflexive Banach spaces).

Proof of Proposition 3.5.1. Let $x(\cdot) \in \mathscr{E}$. Clearly, if S_β is sufficient, then $E_P(x(\cdot) \mid S_\beta)$ does not depend on P. Therefore we can write $E_P(x(\cdot) \mid S_\beta) = E(x(\cdot) \mid S_\beta)$. Since $E_P x(\cdot) = E_P E(x(\cdot) \mid S_\beta)$, implication (3.13) yields

$$E_P\langle x(\cdot), y_P(\cdot)\rangle = E_P\langle E(x(\cdot) \mid S_\beta), y_P(\cdot)\rangle.$$

Thus, by (3.18)

$$E_P\langle x(\cdot), y_P(\cdot)\rangle = E_P\langle x(\cdot), E_P(y_P(\cdot) \mid S_\beta)\rangle$$

for every $x(\cdot) \in \mathscr{E}$. Taking x from a countable dense subset of X, $A \in \mathscr{A}$ and putting $x(\cdot) = x \cdot \chi_A(\cdot)$, we obtain that $y_P(\cdot) = E(y_P(\cdot) \mid S_\beta)$ $\mathscr{P}$-a.e. This means that for every sufficient and $\mathscr{P}$-complete σ-algebra S_β, for every $P \in \mathscr{P}$ and for every Borel set B in Y, $y_P^{-1}(B) \in S_\beta$. Hence $y_P^{-1}(B) \in S$.

PROPOSITION 3.6. *Suppose there exists a family S_P, $P \in \mathscr{P}$, of sub-σ-algebras of $\mathscr{A}$ such that $E_P(x(\cdot) \mid S_P) \equiv^{\mathscr{P}} 0$ for every unbiased estimator of zero in $\mathscr{E}$. Then* (3.13) *holds for each family of functions $y_P(\cdot) \in L_{\Psi,P}$ such that $y_P(\cdot)$ are S_P measurable, $P \in \mathscr{P}$.*

Proof. $E_P(x(\cdot) \mid S_P) \equiv 0$ for all unbiased estimators of zero in $\mathscr{E}$, and S_P measurability of $y_P(\cdot)$ yields (3.14), which is equivalent to (3.13).

4. EFFICIENCY OF BEST UNBIASED ESTIMATORS

In the formulation of the main results we make use of the following definitions.

DEFINITION. An estimator $x_0(\cdot) \in \mathscr{E}$ is said to be locally best at $P = P_0$ in the class of estimators with the same expectation as $x_0(\cdot)$ (or, for short, unbiased locally best at P_0) if

$$R(x_0, P_0) \leqslant R(x(\cdot), P_0)$$

for each $x(\cdot) \in \mathscr{E}$ such that $E_P x(\cdot) \equiv^{\mathscr{P}} E_P x_0(\cdot)$.

DEFINITION. If $x_0(\cdot)$ is an unbiased locally best estimator at every $P \in \mathscr{P}$, then $x_0(\cdot)$ is said to be uniformly best unbiased.

DEFINITION. An estimator $x_0(\cdot) \in \mathscr{E}$ is said to be locally efficient at $P = P_0$ with respect to a given inequality of Cramér–Rao type if this inequality becomes an equality for $x(\cdot) = x_0(\cdot)$ and $P = P_0$.

DEFINITION. If $x_0(\cdot)$ is locally efficient with respect to a given inequality of Cramér–Rao type at every $P \in \mathscr{P}$, then $x_0(\cdot)$ is said to be uniformly efficient with respect to this inequality.

The theorems given below assert that under some weak assumptions the property of an estimator "to be locally efficient" ("to be uniformly efficient") is equivalent to the property "to be locally best unbiased" ("to be uniformly best unbiased").

THEOREM 1. *Let* $x_0(\cdot) \in \operatorname{int} \operatorname{dom} R(\cdot, P_0) \cap \mathscr{E}$. *Then* $x_0(\cdot)$ *is a locally best unbiased estimator at* $P = P_0$ *if and only if there exists an inequality of Cramér–Rao type such that* $x_0(\cdot)$ *is efficient with respect to this inequality at* $P = P_0$.

THEOREM 2. *Let* $x_0(\cdot) \in \bigcap_{P\in\mathscr{P}} \operatorname{int} \operatorname{dom} R(\cdot, P) \cap \mathscr{E}$. *Then* $x_0(\cdot)$ *is a uniformly best unbiased estimator if and only if there exists an inequality of Cramér–Rao type such that* $x_0(\cdot)$ *is uniformly efficient with respect to this inequality.*

In the above "int" stands for interior and is understood as interior in the norm topology of Orlicz spaces $L_{\Phi,P}$.

It is interesting that if the Δ_2-condition

$$\Phi_P(2x, t) \leqslant K_P \Phi_P(x, t) + h_P(t), \qquad h_P(\cdot) \in L_{1,P},$$

is satisfied (cf. Section 2.4), then $\operatorname{dom} R(\cdot, P) = L_{\Phi,P}$. Thus, if the Δ_2-condition holds, then the condition $x_0(\cdot) \in \operatorname{int} \operatorname{dom} T(\cdot, P_0) \cap \mathscr{E}$ in Theorem 1 and the condition $x_0(\cdot) \in \bigcap_{P\in\mathscr{P}} \operatorname{int} \operatorname{dom} R(\cdot, P) \cap \mathscr{E}$ in Theorem 2 reduce simply to $x(\cdot) \in \mathscr{E}$.

As well-known Condition Δ_2 is satisfied, e.g., when the loss function is of the form

$$L(x, t, P) = c(P) \| x - g(P) \|^p, \qquad p \in [1, \infty),$$

where $\| \cdot \|$ stands for a norm in X, $g(P)$ is a function from $\mathscr{P}$ into X and $c(P)$ is a positive real valued function on $\mathscr{P}$.

Theorems 1 and 2 show that efficiency does not characterize estimators of a particular form nor particular families of probability measures, but is closely related to the existence of best unbiased estimators.

The following lemma is a simple consequence of convex analysis (cf. Section

2.7 and [8]), is useful in the proof of Theorems 1 and 2, and extends the Lehmann–Scheffé lemma.

LEMMA 4.1. *Let $L(x, t, P)$ be a normal convex integrand on $X \times T$ and let $R(\cdot, P)$ given by* (3.2) *be a convex functional from $L_{\Phi,P}$ into $(-\infty, +\infty]$. If N is a linear manifold in $L_{\Phi,P}$ and $R(\cdot, P)$ is finite and continuous at $x_0(\cdot) \in N$, then $R(\cdot, P)$ attains at x_0 its infimum over N if and only if there exists a $y_P(\cdot) \in D_{L,P}(x_0(\cdot))$ such that for every $x(\cdot) \in N$*

$$E_P\langle x(\cdot) - x_0(\cdot), y_P(\cdot)\rangle = 0. \tag{4.1}$$

Proof. If $y_P(\cdot) \in D_{L,P}(x_0(\cdot))$ (see Section 2 for the notation) and if (4.1) holds for every $x(\cdot) \in N$, then, by (2.6), the convex functional $R(\cdot, P)$ attains at $x_0(\cdot)$ its infimum over N.

If $R(\cdot, P)$ attains at $x_0(\cdot)$ its infimum, then the theorem stated in Section 2.7 implies the existence of $\varphi_P \in \partial R(\cdot, P)$ $(\partial R(\cdot, P) \subset L'_{\Phi,P})$ such that relation

$$\langle x(\cdot) - x_0(\cdot), \varphi_P\rangle = 0$$

holds for each $x(\cdot) \in N$. The continuity of $R(\cdot, P)$ at $x_0(\cdot)$ yields $x_0(\cdot) \in \operatorname{int} \operatorname{dom} R(\cdot, P)$. Thus, $K_{L,P}(x_0(\cdot)) = 0$ and, by (2.7), $\varphi_P \in D_{L,P}(x_0(\cdot))$. This means that there is a function $y_P(\cdot)$, $y_P(\cdot) \in L_{\Psi,P}$, such that $y_P(t) \in \partial f(x_0(t), t)$ P a.e. and (4.1) holds.

Proof of Theorem 1. If $y_P(\cdot) \in L_{\Phi,P}$, $P \in \mathscr{P}$ and if $y_P(\cdot)$ yields an inequality of Cramér–Rao type, then this inequality is of the form given by (3.15). Let N stand for a manifold of estimators $x(\cdot) \in \mathscr{E}$ with the same expectation as $x_0(\cdot)$. Thus, if $x_0(\cdot)$ is efficient at P_0, $R(\cdot, P_0)$ attains at $x_0(\cdot)$ its infimum over N. Conversely, if $x_0(\cdot) \in \operatorname{int} \operatorname{dom} R(\cdot, P_0) \cap \mathscr{E}$ and if $x_0(\cdot)$ is the unbiased estimator that is locally best at P_0, then Lemma 4.1 yields the existence of $y_P(\cdot) \in L_{\Psi,P}$ such that (4.1) holds for every $x(\cdot) \in N$. Notice that $R(\cdot, P)$ is continuous at $x_0(\cdot) \in \operatorname{int} \operatorname{dom} R(\cdot, P_0) \cap \mathscr{E}$ because every convex function on a locally convex metrizable space is continuous on the interior of its effective domain. It is clear that a fulfilment of condition (4.1) by each $x(\cdot) \in N$ is equivalent to condition (3.14), but only for $P = P_0$. If $P \neq P_0$ we can put $y_P(t) \equiv 0$, e.g., such a family of functions $y_P(\cdot) \in L_{\Psi,P}$, $P \in \mathscr{P}$, satisfies condition (3.14) for every $P \in \mathscr{P}$ and hence leads to the inequality of Cramér–Rao type (3.11). By Proposition 3.1 Orlicz spaces L_{Φ,P_0} and L_{Ψ,P_0} form a dual pair. By Proposition 3.2 $R(\cdot, P_0)$ and $I_M(\cdot, P_0)$ are conjugate to each other. Lemma 4.1, (2.7), and (2.6) yield $y_P(\cdot) \in \partial R(x_0(\cdot), P_0)$. Thus, by (2.8), we have an equality in (4.2) at $P = P_0$. This means that $x_0(\cdot)$ is efficient at $P = P_0$ with respect to a Cramér–Rao type inequality.

Proof of Theorem 2. To prove Theorem 2 it suffices to repeat for every $P \in \mathscr{P}$ the argumentation given in the proof of Theorem 1. The only change is that if $x_0(\cdot)$ is the uniformly best unbiased estimator, then all functions $y_P(\cdot)$ are determined by Lemma 4.1. Hence we obtain a Cramér–Rao type inequality such that $x_0(\cdot)$ is uniformly efficient with respect to it.

Remark 3. It seems to be worthwhile to illustrate (e.g.) Theorem 2 in the simplest case $X = Y = \mathbf{R}$ when a quadratic loss function is used. In this case $L(x, t, P) = \frac{1}{2}(x - g(P))^2$, $M(y, t, P) = \frac{1}{2}y^2 + g(P) \cdot y$, $\Phi_P(x, t) = \frac{1}{2}x^2 + |g(P) \cdot x|$.

As is easy to see, $L_{\Phi,P} = L_{2,P} = L_{\Psi,P}$ and $\mathscr{E} = \bigcap_{P \in \mathscr{P}} \mathscr{L}_{2,P}$. Moreover, $\partial L(x, t, P) = x - g(P)$. Hence, if $x_0(t)$ is a uniformly best unbiased estimator, then by the Lehmann–Scheffé lemma

$$E_P x_0(\cdot) \cdot x(\cdot) = 0$$

for every unbiased estimator of zero $x(\cdot) \in \mathscr{E}$. Thus, if we put $y_P(\cdot) = x_0(\cdot) - g(P)$, then such a family of functions leads to an inequality of Cramér–Rao type. In this case inequality (3.11) has the form

$$\begin{aligned} R(x(\cdot), P) &\geqslant E_P x(\cdot) \cdot (x_0(\cdot) - g(P)) - \tfrac{1}{2}E_P(x_0(\cdot) - g(P))^2 \\ &\quad + g(P)\, E_P(x_0(\cdot) - g(P)) \\ &= E_P(x(\cdot) - g(P))(x_0(\cdot) - g(P)) - \tfrac{1}{2}E_P(x_0(\cdot) - g(P))^2. \end{aligned} \tag{4.2}$$

Now, let $y_P(\cdot) = C_P(x_0(\cdot) - g(P))$. Then

$$R(x(\cdot), P) \geqslant C_P E_P(x(\cdot) - g(P))(x_0(\cdot) - g(P)) - (C_P{}^2/2)\, E_P(x_0(\cdot) - g(P))^2$$

and letting $C_P = E_P(x(\cdot) - g(P))(x_0(\cdot) - g(P))/E_P(x_0(\cdot) - g(P))^2$ we obtain a Cramér–Rao type inequality in the form given by formula (3.17)

$$R(x(\cdot), P) \geqslant \frac{(E_P(x(\cdot) - g(P))(x_0(\cdot) - g(P)))^2}{2E_P(x_0(\cdot) - g(P))^2}. \tag{4.3}$$

Clearly the lower bound for the risk function given by (4.2) is greater or equal to the lower bound given by (4.3). However, in the case $x(\cdot) = x_0(\cdot)$ both (4.2) and (4.3) become equalities. In the general case of $y_P(\cdot)$ satisfying (3.13), inequalities (4.2) and (4.3) can be rewritten in the form

$$R(x(\cdot), P) \geqslant E_P(x(\cdot) - g(P))\, y_P(\cdot) - \tfrac{1}{2}E_P\, y_P(\cdot)^2 \tag{4.2'}$$

and

$$R(x(\cdot), P) \geqslant \frac{(E_P(x(\cdot) - g(P))\, y_P(\cdot))^2}{2E_P\, y_P(\cdot)^2}. \tag{4.3'}$$

Finally, let us note that (4.2) and (4.3) give the inequalities with respect to which the best unbiased estimator $x_0(\cdot)$ is efficient.

Note added in proof. A generalized Jensen's inequality for conditional expectations of Bochner integrable functions and continuous vector-valued convex functions has been proved in a recent paper by Ting On To and Yip Kai Wing, *Pacific J. Math.* **58**, 255–259 (1975). J. Pfanzagl proved a Jensen's inequality in the finite-dimensional case without the continuity assumption on convex function (see *Ann. Prob.* **2**, 490–494 (1974)).

References

[1] Bismut, J. M. (1973). Intégrales convexes et probabilités. *J. Math. Anal. Appl.* **42** 639–673.

[2] Blyth, C. R. (1974). Necessary and sufficient conditions for inequalities of Cramér–Rao type. *Ann. Statist.* **2** 464–473.

[3] Blyth, C. R. and Roberts, D. M. (1972). On inequalities of Cramér–Rao type and admissibility proofs. *In Proceedings of the Sixth Berkeley Symposium on Math. Statist. Prob.* **1** 17–30.

[4] Brønsted, A. (1964). Conjugate convex functions in topological vector spaces. *Math.-Fys. Medd. Danske Vid. Selsk.* **34** 3–27.

[5] Holevo, A. S. (1973). On a generalization of Rao–Cramér inequality. *Theory Prob. Appl.* **18** 371–375 (in Russian).

[6] Isii, K. (1964). Inequalities of the types of Chebyshev and Cramér–Rao and mathematical programming. *Ann. Inst. Statist. Math.* **16** 277–293.

[7] Kozek, A. (1976). Orlicz spaces of functions with values in Banach spaces. *Comment. Math. Prace Mat.* **19** 258–288.

[8] Kozek, A. (1974). On the theory of estimation with convex loss functions. In *Proceedings of the Symposium to Honour Jerzy Neyman.* Warszawa, PWN.

[9] Kozek, A. (1976). "Convex Integral Functionals on Orlicz Spaces." Preprint 89 of Polish Academy of Sciences (submitted to *Comment. Math. Prace Mat.*).

[10] Kozek, A. (1976). On efficiency of best unbiased estimators for convex loss functions. *Bull. Acad. Polon. Sci.* **24** 635–638.

[11] Krasnosielskii, M. A. and Rutickii, Ya. B. (1961). *Convex Functions and Orlicz Spaces.* Noordhoff, Groningen.

[12] Levin, V. L. (1975). Convex integral functionals and the theory of lifting. *Uspekhi Mat. Nauk.* **30** 115–178.

[13] Magiera, R. (1974). On the inequality of Cramér–Rao type in sequential estimation theory. *Zastos. Mat.* **14** 227–235.

[14] Moreau, J. J. (1966–1967). Fonctionnelles convexes. Mimeographed lecture notes. Séminaire sur les Équations aux Derivées Partielles, Collége de France.

[15] Rieffel, M. A. (1968). The Radon–Nikodym theorem for the Bochner integral. *Trans. Amer. Math. Soc.* **131** 466–487.

[16] Rockafellar, R. T. (1971). Integrals which are convex functionals II. *Pacific J. Math.* **39** 439–469.

[17] Rockafellar, R. T. (1971). Convex integral functionals and duality. In *Contributions to Nonlinear Functional Analysis*, pp. 215–236. Academic Press, New York/London.

[18] Suchanecki, Z. On an extension of Jensen's inequality for conditional expectation (unpublished).

[19] Trybuła, S. (1968). Sequential estimation in processes with independent increments. *Dissert. Math.* **60**.

[20] Zacks, S. (1971). *The Theory of Statistical Inference.* Wiley, New York.